Supplementary Text 1: A step-to-step procedure for the Neural Network (NNET) model building

Step 1: After starting the software Rstudio (version 1.1.383. R language version: 3.5.0), the discovery cohort (883 cases) with three iron biomarker expression values were imported by "read.table" function as a data matrix. The data matrix presented with rows as cases and column as three iron biomarker values. Case types were marked as "TB" and "Non-TB" (included HC, PN and RxTB).

Packages were loaded including model building package "caret", NNET package "nnet", ROC curve packages "pROC" and "ROCR", Confidence Interval calculation package "epiR".

Step 2: In order to generate reproducible results, we used a random number generation with "set.seed(20)" function. Before model building, the discovery cohort values were processed with log2 transformation, and then divided randomly at a 3:1 ratio by "createDataPartition" function. Data from the larger (3/4) subset were used for modeling (training set), whereas data from the smaller subset (1/4) were used as the test set. Using "trainControl" function, control parameters were set as 10-fold cross validation and a 5-time-repeat with automatically tuning NNET parameters ("decay" and network size). Finally, the NNET was chosen a 3-5-1 network architecture with decay 0.1 and 26 weights.

Setp 3: Probability of each case ranging from 0 to 1 in training set and test set was

obtained by “extractProb” function using the final NNET model, and that in validation cohort was using “predict” function. Then ROC curve was generated by “roc” function and 95% CI of AUC by “ci.auc” function in “pROC” package according to probabilities and case types. The optical threshold was determined by “coords” function with Youden’s index. Accuracy, sensitivity, specificity and their 95% CI were accessed by “confusionMatrix” and “epi.tests” functions respectively. Calibration of the models was measured by the construction of calibration curves and computation of the Hosmer-Lemeshow goodness-of-fit chi-square statistic. All results were saved into .rda files.

Supplementary Table 1: Clinical and demographic characteristics of involved participants.

Group	Age (years)	Male (%)	Iron (μmol/L)	Ferritin (pmol/L)	Transferrin (g/L)
First group					
HC (n=200)	30.22±10.01	61.50 (123)	20.96±7.60	398.83±269.35	2.46±0.45
TB (n=316) [a,b]	33.36±14.16	63.61 (201)	11.01±7.36	555.95±588.00	1.69±0.49
LTBI (n=167)	31.62±12.31	62.28 (104)	20.17±7.51	389.17±259.53	2.41±0.51
RxTB (n=100)	38.68±12.84	70.00 (70)	23.41±7.24	415.32±381.77	2.65±0.51
PN (n=100)	38.94±13.39	72.00 (72)	19.93±7.61	747.54±725.40	2.04±0.55
Second group					
Non-TB (n=148)[c]	32.89±12.68	57.43 (85)	19.35±7.08	470.52±485.31	2.42±0.49
Definite TB (n=653)	40.96±17.46	65.85 (430)	12.27±7.83	852.37±850.85	1.71±0.50
Probable TB (n=131)	37.75±16.70	62.60 (82)	12.39±6.72	864.31±742.52	1.68±0.42
Questionable (n=12)	46.83±14.11	66.67 (8)	12.33±6.47	788.32±623.48	1.91±0.34

Values are reported as the means ± standard error of the mean, or % (n).

[a]All definite TB.

[b]45 patients were followed up for 1 month.

[c]Includes 115 with HC, 26 with PN, two with asthma, two with bronchiectasis, one with COPD, one with lung cancer and one with pulmonary silicosis.

Abbreviations: TB, tuberculosis; HC, healthy control; LTBI, latent tuberculosis infection; RxTB, recovered tuberculosis; PN, pneumonia; COPD, chronic obstructive pulmonary disease.

Supplementary Table 2: Diagnostic efficiency of different models in diagnosing TB in the discovery cohort

Parameter	Training set (N=663: 152 HC, 237 TB, 132 LTBI, 63 RxTB, 79 PN)				Test set (N=220: 48 HC, 79 TB, 35 LTBI, 37 RxTB, 21 PN)			
	NNET	SVM	LDA	rPart	NNET	SVM	LDA	rPart
Kappa	0.68	0.69	0.64	0.66	0.64	0.57	0.64	0.56
AUC	0.91 (0.89, 0.94)	0.91 (0.88, 0.93)	0.90 (0.87, 0.92)	0.87 (0.84, 0.90)	0.90 (0.86, 0.94)	0.86 (0.81, 0.92)	0.90 (0.86, 0.94)	0.83 (0.78, 0.89)
Accuracy	0.85(0.82, 0.88)	0.85 (0.82, 0.88)	0.83 (0.80, 0.86)	0.84 (0.81, 0.87)	0.84 (0.79, 0.89)	0.81 (0.75, 0.86)	0.84 (0.78, 0.88)	0.80 (0.74, 0.85)
Sensitivity	0.83 (0.77, 0.87)	0.86 (0.81, 0.90)	0.83 (0.77, 0.87)	0.80 (0.75, 0.85)	0.70 (0.58, 0.79)	0.70 (0.58, 0.79)	0.72 (0.61, 0.82)	0.68 (0.57, 0.78)
Specificity	0.86 (0.83, 0.89)	0.85 (0.81, 0.88)	0.83 (0.79, 0.87)	0.86 (0.83, 0.89)	0.92 (0.86, 0.96)	0.87 (0.80, 0.92)	0.90 (0.84, 0.94)	0.87 (0.80, 0.92)
PPV	0.77 (0.71, 0.82)	0.76 (0.70, 0.81)	0.73 (0.67, 0.78)	0.77 (0.71, 0.82)	0.83 (0.72, 0.91)	0.74 (0.63, 0.84)	0.80 (0.69, 0.89)	0.74 (0.62, 0.84)
NPV	0.90 (0.87, 0.93)	0.92 (0.88, 0.94)	0.90 (0.86, 0.92)	0.89 (0.85, 0.92)	0.84 (0.78, 0.90)	0.84 (0.77, 0.89)	0.85 (0.79, 0.91)	0.83 (0.76, 0.89)
PLR	5.97 (4.68, 7.62)	5.64 (4.48, 7.10)	4.89 (3.93, 6.09)	5.89 (4.60, 7.54)	8.92 (4.97, 16.03)	5.17 (3.32, 8.05)	7.27 (4.34, 12.17)	5.07 (3.25, 7.91)
NLR	0.20 (0.15, 0.27)	0.16 (0.12, 0.23)	0.21 (0.16, 0.28)	0.23 (0.18, 0.30)	0.33 (0.24, 0.46)	0.35 (0.25, 0.49)	0.31 (0.22, 0.44)	0.37 (0.26, 0.51)

Diagnostic efficiencies of four supervised models (LDA, SVM, NNET and rPart) are shown. Data in parentheses represent 95% confidence intervals.

Abbreviations: HC, healthy control; LTBI, latent tuberculosis infection; TB, tuberculosis; RxTB, cured TB; PN, pneumonia; AUC, area under the receiver operating characteristic curve; PPV, positive predictive value; NPV, negative predictive value; PLR, positive likelihood ratio; NLR, negative likelihood ratio; LDA, Linear Discriminant Analysis; SVM, support vector machine; NNET, neural networks; rPart, decision tree.

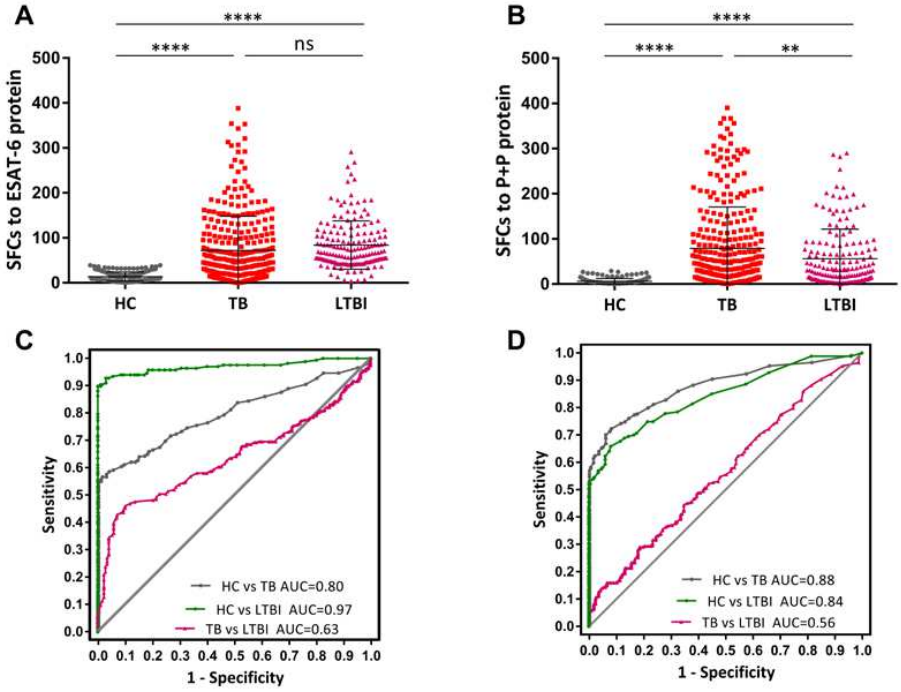
Supplementary Table 3: Key parameters for intergroup comparisons between HC, LTBI, RxTB or PN with TB

Parameter	Controls[a]	HC	LTBI	RxTB	PN
AUC	0.91 (0.89, 0.93)	0.93 (0.91, 0.95)	0.91 (0.88, 0.93)	0.95 (0.93, 0.97)	0.83 (0.79, 0.88)
Accuracy	0.84 (0.81, 0.87)	0.85 (0.82, 0.88)	0.83 (0.79, 0.86)	0.84 (0.81, 0.88)	0.78 (0.74, 0.82)
Sensitivity	0.82 (0.77, 0.86)	0.82 (0.77, 0.86)	0.82 (0.77, 0.86)	0.82 (0.77, 0.86)	0.82 (0.77, 0.86)
Specificity	0.86 (0.82, 0.88)	0.91 (0.86, 0.95)	0.86 (0.79, 0.91)	0.93 (0.86, 0.97)	0.67 (0.57, 0.76)
PPV	0.76 (0.71, 0.80)	0.93 (0.90, 0.96)	0.91 (0.88, 0.94)	0.97 (0.95, 0.99)	0.89 (0.84, 0.92)
NPV	0.89 (0.86, 0.92)	0.76 (0.70, 0.81)	0.71 (0.64, 0.77)	0.62 (0.53, 0.69)	0.54 (0.44, 0.63)
PLR	5.65 (4.59, 6.94)	9.07 (5.82, 14.14)	5.68 (3.91, 8.26)	11.66 (5.70, 23.87)	2.47 (1.86, 3.29)
NLR	0.21 (0.17, 0.27)	0.20 (0.16, 0.26)	0.21 (0.17, 0.27)	0.20 (0.16, 0.25)	0.27 (0.21, 0.36)

[a]Merged data from HC, LTBI, RxTB and PN groups.

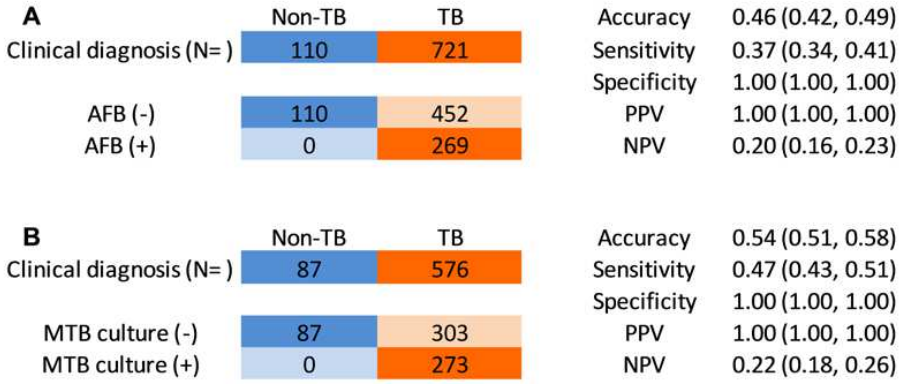
Data in parentheses represent 95% confidence intervals.

Abbreviations: TB, tuberculosis, HC, healthy control; LTBI, latent tuberculosis infection; RxTB, cured TB; PN, pneumonia; PPV, positive predictive value; NPV, negative predictive value; PLR, positive likelihood ratio; NLR, negative likelihood ratio.



Supplementary Figure 1: ROC analysis of T-SPOT results between TB (n=316), HC (n=200) and LTBI (n=167). The numbers of interferon-γ SFCs in response to MTB-specific antigens ESAT-6 (A, C) and CFP-10 (B, D) were compared by one-way analysis of variance and Tukey's multiple comparison test. All cases of TB in the first group are definite TB. Data are expressed as the means ± standard error of the mean. **P<0.01; **** P<0.0001.

Abbreviations: TB, tuberculosis; HC, healthy control; LTBI, latent tuberculosis infection; SFC, spot forming cell; MTB, *Mycobacterium tuberculosis*; ns, not significant; ROC, receiver operating characteristic curve; AUC, area under the ROC.



Supplementary Figure 2: Diagnostic performance of AFB and MTB cultures in an independent validation cohort. In the validation cohort, 89.2% (831/932) and 71.1% (663/932) were taken forward for AFB and MTB culture, respectively. The numerical agreement with the clinical diagnosis and key diagnostic parameters of AFB (A) and MTB cultures (B) are listed.

Abbreviations: TB, tuberculosis; AFB, acid fast bacilli staining; MTB, *Mycobacterium tuberculosis*; PPV, positive predictive value; NPV, negative predictive value.